THE AVERAGING HOMOTOPY

F.Chaplais
Centre Automatique et Systèmes
École Nationale Supérieure des Mines de Paris
35 rue Saint Honoré
77305 Fontainebleau Cedex FRANCE

Abstract

Classical averaging is an asymptotic theory in the sense that it considers the problem of finding a limit for a sequence of differential systems. We present here another formulation that is based on an homotopy between any two vectorfields. This formulation accounts both for classical averaging and for regular perturbations. It can also be used to justify the use of numerical windowed averaging schemes without any asymptotic imbedding.

1 Introduction

Averaging is classically presented as a technique to asymptotically approximate the solutions of an ε-dependent ordinary differential equation

$$\frac{dx_\varepsilon}{dt} = f(x_\varepsilon, t, \tau)_{\tau=\frac{t}{\varepsilon}} \quad , \quad x_\varepsilon(0) = x_0 \tag{1}$$

by that of another, supposedly simpler system:

$$\frac{dy}{dt} = \bar{f}(y, t) \quad , \quad x_\varepsilon(0) = x_0 \tag{2}$$

as ε tends to zero [1] .

One important feature of equation (1) is that the limit of $\frac{t}{\varepsilon}$ is singular in ε as ε goes to zero. Generally (see Sanderst-Verhulst[5] for a comprehensive monograph on the subject), the regularity of the solutions x with respect to ε is recovered by assuming that f is a so-called *Krilov-Bogolioubov-Mitropolski (KBM) vectorfield* with average $\bar{f}$, e.g., f satisfies the following ergodicity assumption:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T (f(x,t,\tau) - \bar{f}(x,t,\tau))d\tau = 0 \tag{3}$$

[1] This is a "normal+fast time" formulation of averaging. Notice that, in these settings, the "slow" time equivalent of (1) is already in normal form, e.g. there is no singular perturbation in (1).

This includes the case when f is periodic in τ with a period independent of x; $\bar{f}$ is then the average of f as usually defined for periodic functions. Eckhaus[3] refined the concept further by introducing local averages and discriminating the order of the time scale from the order of the horizon T.

The ergodicity assumption (3) is quite natural when the perturbed system (1) is the rewriting on a fast time scale of some regularly perturbed differential system (in its normal form), such as the Van Der Pol equation or many other examples[5]. However, in practice, there are many cases where the unstationnarity of (1) does not originate in some time rescaling, but is simply given as raw data[2]. Yet, there is a strong feeling that averaging techniques could be still used in this case.

A typical case of this happens in the study of the input-output relationship in control theory. The unstationnarity of the system is expressed there in terms of some input signal u which corresponds either to the action of some user, or to the influence of a (non-controlled) perturbation. The system is written as

$$\frac{dx}{dt} = f(x(t), u(t)) \tag{4}$$

where the function u denotes the *input* of the system. When u stands for a perturbation, e.g. a signal that is neither controlled nor precisely known, it is interesting to evaluate the influence of this perturbation on the trajectory, even when *a priori* knowledge on it is restricted to some bounds on its amplitude or its frequency distribution. We can see on this example that no ergodicity assumption can be made on the perturbation, since the information on it is reduced to the minimum.

However, the idea remains that inputs of small amplitude lead to small deviations of the trajectory (this may be considered as a case of regular perturbations), and that inputs in high frequency domains produce also small deviations (and this may be seen as a case of averaging; in signal processing terms, we would say that strictly causal systems are low pass filters).

The purpose of the paper is twofold

- present averaging as a case of *regular* perturbations, or, specifically, an homotopy between two dynamics
- show that using suitable signal processing tools to define the "average" dynamics is an effective means of have close trajectories when comparing the original system to the processed system.

The article is organized as follows. Section 2 presents the averaging homotopy and how if is related to classical averaging. Section 3 investigate the difference between the trajectory analysis by using first order estimates. Section 4 shows a constructive example that defines an averaging homotopy, by using local averages of the system. Section 5 discusses higher order expansions with respect to the hopotopy parameter ϵ; in particular, theorems 5.4, 5.6 and 5.7 investigate the converge of the series for $\epsilon = 1$.

[2] in particular, there is actually no small parameter ε in the equation from the start.

2 The averaging homotopy

Let us rewrite (1) under its integral form

$$x_\varepsilon(t) = x_0 + \int_0^t f(x_\varepsilon(s), s, \frac{s}{\varepsilon})ds \tag{5}$$

After some integration by part, it becomes

$$\begin{aligned} & x_\varepsilon(t) - \varepsilon\Delta(x_\varepsilon(t), t, \frac{t}{\varepsilon}) \\ & \qquad = x_0 + \int_0^t \bar{f}(x_\varepsilon(s), s)ds - \varepsilon \int_0^t \left[\frac{\partial \Delta}{\partial x} f + \frac{\partial \Delta}{\partial t}\right](x_\varepsilon(s), s, \frac{s}{\varepsilon})ds \end{aligned} \tag{6}$$

where we define Δ to be

$$\Delta(x, t, \tau) = \int_0^\tau \left[f - \bar{f}\right](x, t, \theta)d\theta \tag{7}$$

If $\Delta(x, t, \tau)$ remains bounded for all τ, as in the periodic case, we can see that, at least locally, x_ε converges to y when ε goes to 0. This result, as well as the transformation (6), is quite classical, and can be found in [5].

Lets us now assume that f is not given as a two time scale system or, in other words, that we have $\varepsilon = 1$. Then we see that it is enough to have Δ and its derivative to be small for y to approximate x_ε. This is true for any $\bar{f}$. By analogy to the case of regular perturbations, where one transforms a vector field into a vectorfield g by linear interpolation, we use the previous remark to define the averaging homotopy for *any* two vectorfields f and $\bar{f}$:

Definition 2.1 (Averaging homotopy) *Let f and $\bar{f}$:*

$$\begin{array}{rcl} f : \mathbb{R}^n \times \mathbb{R} & \longrightarrow & \mathbb{R}^n \\ (x, t) & \longmapsto & f(x, t) \end{array} \tag{8}$$

$$\begin{array}{rcl} \bar{f} : \mathbb{R}^n \times \mathbb{R} & \longrightarrow & \mathbb{R}^n \\ (x, t) & \longmapsto & \bar{f}(x, t) \end{array} \tag{9}$$

that we shall assume to be as smooth as necessary. Define $\Delta(x, t)$ by

$$\Delta(x, t) = \int_0^t \left[f - \bar{f}\right](x, s)ds \tag{10}$$

We shall assume that $\frac{\partial \Delta}{\partial x}$ is smaller than the identity in a neighborhood $\mathcal{V}$ of $(x_0, 0)$. In this neighborhood, we define the averaging deformation of $\bar{f}$ into f by

$$\begin{array}{rcl} AD : [0, 1] & \longrightarrow & (\mathbb{R}^n)^{\mathcal{V}} \\ \varepsilon & \longmapsto & f_\varepsilon(x, t) = \left(Id - \varepsilon \frac{\partial \Delta}{\partial x}\right)^{-1} \left[\bar{f} + \varepsilon(f - \bar{f}) - \varepsilon \frac{\partial \Delta}{\partial x} f\right](x, t) \end{array} \tag{11}$$

and we define $x(\varepsilon, t)$ as the maximum solution of

$$\frac{dx}{dt} = f_\varepsilon(x, t) \quad , \quad x(0) = x_0 \tag{12}$$

Comments

- the inverse in equation (11) is a matrix inverse
- we check that, for $\varepsilon = 0$, $f_0 = \bar{f}$, and that, for $\varepsilon = 1$, $f_1 = f$
- The state x_ε satisfies

$$x_\varepsilon(t) - \varepsilon\Delta(x_\varepsilon(t), t) = \\ x_0 + \int_0^t \bar{f}(x_\varepsilon(s), s)ds - \varepsilon \int_0^t \left[\frac{\partial \Delta}{\partial x} f\right](x_\varepsilon(s), s)ds \quad (13)$$

(actually, the definition of x_ε was obtained by differentiating the previous equation, which is formally similar to (6)). This shows that having a small Δ is enough to have x close to y, something that is not obvious in the differential form (11) (12). If we had substituted Δ to its gradient in the homotopy (11), we might have not been able to recover this result (unless Δ itself was a gradient).

- assume now that f is under the form $f(x, \frac{t}{\mu})$, with f periodic with respect to the last variable; then, if we take $\bar{f}$ to be the average of with respect to this last variable, it is clear that Δ goes to 0 with μ and (13) becomes

$$x_\varepsilon(t) - \mu\varepsilon\Delta(x_\varepsilon(t), \frac{t}{\mu}) = x_0 + \int_0^t \bar{f}(x_\varepsilon(s))ds - \mu\varepsilon \int_0^t \left[\frac{\partial \Delta}{\partial x} f\right](x_\varepsilon(s), \frac{s}{\mu})ds \quad (14)$$

We see that any expansion with respect to ε will also be an asymptotic expansion with respect to μ, even though it may be different from the expansions that are considered in classical averaging. This can be extended, of course, to KBM vectorfields. It will be developed in section 5.

- Δ can be made small (at least on a finite horizon) by simply taking $\bar{f}$ uniformly close to f. This means that we can also handle the case of regular perturbations in this framework, even though it will be with a different homotopy, and thus, different expansions.

Case of a perturbation input Assume now that $f(x, t)$ is of the form $g(x, u(t))$, where u is a perturbation data which cannot be modelled. As a consequence, there is not reason to adopt an asymptotic point of view on the matter. However, if we use the averaging homotopy, we have freedom in choosing the “average” $\bar{f}$. To do so, we can use a low pass filter LP and define $\bar{f}$ as

$$\bar{f}(x, t) = LP(f)$$

where the low pass filter is only applied to time variation of f, x being a static variable. Then Δ is the indefinite integral of $f(x, t)$ through a *high pass* filter, and as such Δ can be small.

3 First order averaging

In this section we are going estimate the distance between x_ε and y in terms of estimates on f and, essentially, Δ. For $\varepsilon = 1$, this gives an estimation on the difference between the solutions of $dx/dt = f$ and of $dx/dt = \bar{f}$.

3.1 Estimates based on Lipschitz constants

Assumption 1 *We assume that f is continuous, $\bar{f}$ locally Lispschitz, and Δ is of class $\mathcal{C}^1$, all of this with respect to x. We consider a neighborhood V of $(x_0, 0)$ in the (x, t) domain on which the following hold:*

$$\begin{aligned} \|\bar{f}(x,t) - \bar{f}(y,t)\| &\leq \bar{F}_1 \|x - y\| && (15) \\ \|\Delta(x,t)\| &\leq D_0 && (16) \\ \|f(x,t)\| &\leq F_0 && (17) \end{aligned}$$

We also assume that "$\frac{\partial \Delta}{\partial x}$ is not too big", that is, Δ satisfies

$$\|\frac{\partial \Delta}{\partial x}(x,t)\| \leq D_1 < 1 \tag{18}$$

on V.

We shall take advantage of the averaging formulation to draw estimates finer than those simply derived from the amplitude of $\bar{f}$.

Theorem 3.1 *As long as $(x(\varepsilon, t), t)$ and $(x(0,t), t) = (y(t), t)$ stay in V, the following holds:*

$$\|x(\varepsilon, t) - x(0,t)\| \leq \varepsilon \left[\frac{e^{\bar{F}_1 t} - 1}{\bar{F}_1} (D_0 + D_1 F_0) - \frac{t}{\bar{F}_1} D_1 F_0 \right] \tag{19}$$

Proof: since we have $D_1 < 1$, $Id - \varepsilon\Delta$ is invertible in x for $\varepsilon \leq 1$. Hence, z and x are defined, at least locally. From (13) taken at ε and at 0, we get

$$\|x(\varepsilon, t) - x(0,t)\| \leq \int_0^t \bar{F}_1 \|x(\varepsilon, s) - x(0, s)\| \, ds + \varepsilon D_0 + \varepsilon \int_0^t D_1 F_0 \, ds \tag{20}$$

which leads to the result. ■

Comments

- for $\varepsilon = 1$, we get an estimation of the difference $x - y$, where x is the solution of $dx/dt = f$, and y is the solution of $dy/dt = \bar{f}$.
- remark that $x(\varepsilon, t) - x(0, t)$, and hence, $x - y$, is small when D_0 and D_1 are small. This can be achieved by having a $\bar{f}$ that is close to f in amplitude (regular perturbations), or by stripping f of its high frequency components (averaging) to get $\bar{f}$.

- if f is μ-periodic in time, and $\bar{f}$ is the usual average, then D_0 and D_1 are of order 1 with respect to μ[3], while F_0 and $\bar{F}_1$ are of order 0. This proves that there exist a constant K such that $\|x(1,t) - x(0,t)\| \leq K\mu$, and that we can recover the classical results on periodic averaging.
- this estimate, of course, can be used tell how long x stays in V.

In the next section we assume a little more regularity to make use of the tangent system at $\varepsilon = 0$.

3.2 Estimates based on the first derivative

3.2.1 Equation on the transition matrix

To simplify the computations, we shall define g by

$$g = \frac{\partial \Delta}{\partial x} f.$$

We assume now that f and $\bar{f}$ are $\mathcal{C}^1$. Instead of computing a Lipschitz constant on x with respect to ε, we are going to use the Taylor formula to estimate $x(\varepsilon) - x(o)$. Let us first compute $\partial x / \partial \varepsilon$. To do that, we differentiate (13) with respect to ε. This gives:

$$\left(Id - \varepsilon \frac{\partial \Delta}{\partial x} \right) \frac{\partial x}{\partial \varepsilon} - \Delta = \int_0^t \left[\left(\frac{\partial \bar{f}}{\partial x} - \varepsilon \frac{\partial g}{\partial x} \right) \frac{\partial x}{\partial \varepsilon} - g \right] ds \tag{21}$$

where all the functionals are computed at $(x(\varepsilon, t), t)$. If we let $\tilde{x} = (Id - \varepsilon \partial \Delta / \partial x) \partial x / \partial \varepsilon - \Delta$, we may rewrite this as

$$\tilde{x} = \int_0^t \left[\left(\frac{\partial \bar{f}}{\partial x} - \varepsilon \frac{\partial g}{\partial x} \right) \left(Id - \varepsilon \frac{\partial \Delta}{\partial x} \right)^{-1} (\tilde{x} + \Delta) - g \right] ds \tag{22}$$

Let us call Φ_ε the transition matrix associated to the above affine equation. After some reordering with respect to ε, we can see that Φ_ε satisfies

$$\frac{\partial \Phi_\varepsilon}{\partial t} = \left[\frac{\partial \bar{f}}{\partial x} + \varepsilon \left(\frac{\partial \bar{f}}{\partial x} \frac{\partial \Delta}{\partial x} - \frac{\partial g}{\partial x} \right) \left(Id - \varepsilon \frac{\partial \Delta}{\partial x} \right)^{-1} \right] \Phi_\varepsilon \tag{23}$$

Since $\tilde{x}$, and hence, $\frac{\partial x}{\partial \varepsilon}$ [4] is determined by a convolution with Φ_ε, it is interesting to have an idea of the norm of this convolution operator, and of its proximity to the corresponding operator at $\varepsilon = 0$.

[3] this can be seen as a consequence of theorem 4.1

[4] and, in fact, all the further derivatives of x with respect to ε, as we shall see later.

3.2.2 Estimates on the transition matrix

Lemma 3.1 *Let $\bar{F}_1$ and G_1 some estimates on the first derivatives of $\bar{f}$ and g on a neighborhood $\mathcal{V}$ of the initial conditions. Given a kernel $K(t,s)$, we denote by $\|K\|_*$ the norm of the convolution operator from L^∞ to L^∞ that is associated to it*[5]*, e.g.* $\max_{a,b\,\in[0,T]} \int_a^b \|K(t,s)\| ds$ *and we denote by φ_ε the norm $\|\Phi_\varepsilon\|_*$*[6]*. Then, as long as $(x(\varepsilon,t),t)$ stays in $\mathcal{V}$ for $0 \le \varepsilon \le 1$, one has*

$$\left\| \left(Id - \varepsilon \frac{\partial \Delta}{\partial x} \right)^{-1} \right\| \le \frac{1}{1 - \varepsilon D_1} \tag{24}$$

and, for ε such that $\varepsilon[\varphi_0(\bar{F}_1 D_1 + G_1) + D_1] \le 1$, we have

$$\|\Phi_\varepsilon - \Phi_0\|_* \le \frac{\varepsilon c(\varepsilon) \varphi_0^2}{1 - \varepsilon c(\varepsilon) \varphi_0} \tag{25}$$

where $c(\varepsilon)$ stands for $\frac{\bar{F}_1 D_1 + G_1}{1 - \varepsilon D_1}$.

Proof: The estimate on A^{-1} comes from its expression as a (positive) power series in Δ.

Let us now denote by $C(\varepsilon)$ the matrix $\left(\frac{\partial \bar{f}}{\partial x}\frac{\partial \Delta}{\partial x} - \frac{\partial g}{\partial x}\right)\left(Id - \varepsilon \frac{\partial \Delta}{\partial x}\right)^{-1}$. We have $\|C(\varepsilon)\| \le c(\varepsilon)$. After solving (23), we get on $\Phi_\varepsilon - \Phi_0$:

$$\begin{aligned}
\int_a^b \|\Phi_\varepsilon(b,s) - \Phi_0(b,s)\| ds \quad &\le \quad \varepsilon \int_a^b \int_s^b \|\Phi_0(t,\theta)\| \, \|C(\varepsilon)\| \, \|\Phi_0(t,\theta)\| \; d\theta \\
&+ \quad \varepsilon \int_a^b \int_s^b \|\Phi_0(t,\theta)\| \, \|C(\varepsilon)\| \, \|\Phi_\varepsilon(\theta,s) - \Phi_0(t,\theta)\| \; d\theta \\
&\le \quad \varepsilon c(\varepsilon) \varphi_0^2 + \varepsilon c(\varepsilon) \varphi_0 \|\Phi_\varepsilon - \Phi_0\|_*
\end{aligned}$$

which leads to the result. ∎

Comment The estimate on Φ_ε holds for $\varepsilon \in [0,1]$ if $\varphi_0(\bar{F}_1 D_1 + G_1) \le 1 - D_1$, that is, essentially, if Δ is small and φ_0 is not too large. Notice that $c(\varepsilon)$ is small if Δ is small. Finally, the estimation on $\Phi_\varepsilon - \Phi_0$ can be used to obtain a bound on φ_ε; we get $\varphi_\varepsilon \le \frac{\varphi_0}{1 - \varepsilon c(\varepsilon) \varphi_0}$.

Since φ_0 in may depend on the averaging scheme used to compute $\bar{f}$, it is also interesting to have an estimation on φ_0 with respect to φ_1, which may be part of the actual data of the averaging problem. We have for φ_0 an estimation that is exactly symmetrical of that obtained for φ_1 in the previous lemma:

Lemma 3.2 *The assumptions are that same as in lemma 3.1; moreover, we assume $c(1)\varphi_1 < 1$. Then*

$$\varphi_0 \le \frac{\varphi_1}{1 - c(1)\varphi_1} \tag{26}$$

[5]in control theory, this is the Bounded Input Bounded Output norm. This bound always exists on a finite horizon; if K satisfies $\|K(t,s)\| \le A e^{-\alpha(t-s)}$, we have $\|K\|_* \le \frac{A}{\alpha}$, and the norm is defined for $T = +\infty$.

[6]$\Phi_0(t,s)$ is the transition matrix of $\frac{\partial \bar{f}}{\partial x}(x(0,t),t)$.

Proof: We have

$$\Phi_0(t,s) = \Phi_1(t,s) - \int_s^t \Phi_1(t,\theta)C(1)\Phi_0(\theta,s)ds \tag{27}$$

and the rest of the prof follows the line of that of the previous lemma. ■

From now on, we shall assume that $\varphi_0(\bar{F}_1 D_1 + G_1) \leq 1 - D_1$; along with $D_1 < 1$, these are the only restrictions on the choice of f and $\bar{f}$. As we saw, the can be checked by assuming that Δ is small enough; this, in turn, can be achieved, for instance, by taking a sufficiently small averaging horizon in the scheme described later in section 4. We shall also replace, for all purposes, φ_ε by its estimate $\varphi_0/(1 - \varepsilon c(\varepsilon)\varphi_0)$, which is an increasing function of ε.

3.2.3 Estimates on the trajectories

Theorem 3.2 *Under the previous assumptions, we have, as long as $(x(\varepsilon_1,t),t)$ stays in $\mathcal{V}$ for all $\varepsilon_1 \in [0,\varepsilon]$:*

$$\|x(\varepsilon,t) - x(0,t)\| \leq \varepsilon \frac{\varphi_\varepsilon(c(\varepsilon)D_0 + D_1 F_0) + D_0}{1 - \varepsilon D_1} \tag{28}$$

where D_0 is the sup norm of Δ on V, D_1 is the sup norm of $\frac{\partial \Delta}{\partial x}$ on V, and F_0 is the sup norm of f on V.

Proof: We have

$$\frac{\partial x}{\partial \varepsilon} = \left(Id - \varepsilon \frac{\partial \Delta}{\partial x} \right)^{-1} (\tilde{x} + \Delta) \tag{29}$$

and

$$\tilde{x} = \int_0^t \Phi_\varepsilon(t,s) \left[\left(\frac{\partial \bar{f}}{\partial x} - \varepsilon \frac{\partial g}{\partial x} \right) \left(Id - \varepsilon \frac{\partial \Delta}{\partial x} \right)^{-1} \Delta - g \right] ds \tag{30}$$

so that $\|\tilde{x}\| \leq \varphi_\varepsilon(c(\varepsilon)D_0 + D_1 F_0)$, and $\|\partial x/\partial \varepsilon\| \leq (\|\tilde{x}\| + D_0)/(1 - \varepsilon D_1)$. ■

Comments

- for $\varepsilon = 1$, we get an estimation of the difference $x - y$, where x is the solution of $dx/dt = f$, and y is the solution of $dy/dt = \bar{f}$.
- if f is μ-periodic in time and $\bar{f}$ is the usual average, then D_1 and G_1 are of order 1 with respect to μ, while $\bar{F}_1$ is of order 0. This shows that $c(\varepsilon)$ is of order in μ, and hence $\|\Phi_\varepsilon - \Phi_0\|_*$; φ_ε is thus of order 0. Since D_0, as we said before, is of order 1, we see that the right handside of (28) is of order 1 with respect to μ.
- if we have $\varphi_0(\bar{F}_1 D_1 + G_1) \leq 1 - D_1$, then we can replace φ_ε by $\frac{\varphi_0}{1-\varepsilon c(\varepsilon)\varphi_0}$. In this case, we have estimates that are increasing functions of ε for $\varepsilon \in [0,1]$. Let us call B the value of this bound for $\varepsilon = 1$. If $\mathcal{V}$ is then defined by $\|x - x_0\| \leq R$ and $0 \leq t \leq T$, then $(x(\varepsilon,t),t)$ stays in V as long as

$\|x(0,t)-x_0\|+B \leq R$ and $0 \leq t \leq T$. This implies, of course, that we have $B \leq R$, something which can be achieved by having a small Δ (or a small μ in the periodic case). We shall refer to this (sufficient) condition as condition $\mathcal{V}_R$.

4 An example of averaging synthesis

We have seen that the quality of the approximations, whether in (19) or in (28), is essentially determined by the amplitude of Δ and its derivatives. In the periodic case, we have seen that the choice of the usual average for $\bar{f}$ in the case of small periods is a good choice in the sense that Δ is then small. In the general case, the question remains of the determination of the "average" $\bar{f}$ in order to have a "small" Δ. A trivial choice consists in taking $\bar{f}=f$. Another choice consists in taking for $\bar{f}$ a regular perturbation of f. For instance, a Taylor expansion of f around a reference trajectory would be a (locally) suitable choice; this case will be considered in the next section. However, this choice does not correspond to the usual averaging setup, even though it may (and should) be considered.

In this section, we show that a simple windowed average scheme is enough to provide a suitable $\bar{f}$ in all cases.

We consider here a function $f(x,t)$ on $\mathbb{R}^n \times \mathbb{R}$, essentially locally bounded in t and $d>0$. We define the windowed average $\bar{f}$ by

$$\bar{f}(x,t)=\frac{1}{d}\int_{k\delta}^{(k+1)d} f(x,s)ds \quad \text{for} \quad t\in[k\delta,(k+1)d[\tag{31}$$

The following theorem proves that, for a given f, an arbitrarily small Δ can be obtained by choosing a sufficiently small step d.

Theorem 4.1 (Bound on Δ from the standard deviation) *Let $\Delta(x,t)=\int_0^t (f(x,s)-\bar{f}(x,s))ds$. Define*

$$\sigma_k(f)=\sqrt{\frac{1}{d}\int_{k\delta}^{(k+1)d}\|f(x,s)-\bar{f}(x,s)\|^2 ds} \tag{32}$$

and $\sigma(f)=\max \sigma_k$. Then

$$\|\Delta(x,\bullet)\|_{L^\infty([0,T])} \leq d\sigma(f) \leq d\|f(x,\bullet)\|_{L^\infty([0,T])} \tag{33}$$

Proof: we shall omit to mention x in the proof, since it is only a fixed parameter. Since the integral of f coincides with that of its average $\bar{f}$ on all the $[k\delta,(k+1)d]$'s, we have, for $t\in[N\delta,(N+1)d[$:

$$\begin{aligned}
\|\Delta(t)\| &= \|\int_{N\delta}^{t}(f(s)-\bar{f}(s))ds\| \\
&\leq \sqrt{d}\|f-\bar{f}\|_{L^2([N\delta,(N+1)d])} \\
&= d\sigma_N(f) \\
&\leq d\|f\|_{L^\infty([N\delta,(N+1)d])}
\end{aligned}$$

since, for convexity reasons, we always have $\sigma(f) \leq \|f\|_\infty$. ■

Comments

- this result can be extended to the derivatives of Δ by replacing f by its corresponding derivatives.
- if f is d-periodic, then $\bar{f}$ is constant and equal to the usual average, and we recover the classical results of periodic averaging.
- for obvious reasons, we did not want to use bounds on $\partial f/\partial t$ in this estimation. However, if f in q-Hölder of order q in t with constant λ, then $\sigma(f) \leq \lambda d^q$. For continuous functions, $\sigma(f)$ goes to zero with d (uniformly on compact subsets).

Even though this fixed-width windowed average scheme may be not the smartest one in all cases, it is interesting for several reasons: it is very simple, it does kill the higher frequencies of f, and, in particular, it includes the classical averaging process for periodic functions, and it works in all cases. However, it is quite possible that, in cases where the nature of dynamics vary from time to time, from slowly varying to stiff or to fast oscillating, for instance, that a variable step size scheme would prove to be more efficient.

5 Higher order approximations

The purpose of this section is to use the averaging homotopy to derive an expansion of $x(\varepsilon, t)$ with respect to ε, and to prove that this expansion is also an expansion with respect to Δ, to the period μ in the periodic case, and the step d in the case of the windowed averaging scheme of section 4. To achieve that, we compute the derivatives of x with respect to ε; then we compute some estimates on this derivatives to obtain a bound on the error on x_ε when we replace it by its Taylor expansion at $\varepsilon = 0$. Finally, we bound the estimates themselves in the case of periodic vectorfields of in the case of windowed averaging to prove our claim that this Taylor expansion in also a valid expansion with respect to the period μ or the window length d.

In order to have reasonably writable derivatives, we shall use the intermediate variable z defined by

$$z(\varepsilon, t) = x(\varepsilon, t) - \varepsilon\Delta(x(\varepsilon, t), t) \tag{34}$$

We have:

$$\frac{\partial z(\varepsilon, t)}{\partial t} = \bar{f}(x(\varepsilon, t), t) - \varepsilon \left[\frac{\partial \Delta}{\partial x} f\right]_{(x(\varepsilon,t),t)} \tag{35}$$

5.1 Differentiation of x and z with respect to ε

We shall assume during the rest of this paper that $\frac{\partial\Delta}{\partial x}(x_0,0) < 1$ and that f and $\bar{f}$ are regular enough, so that $x(\varepsilon,t)$ exists and is regular for $0 \le \varepsilon \le 1$ and t small enough.

Let us first recall some results on composite differentiation.

5.1.1 Derivatives of the composition of two functions

If f is a function of x, we shall denote by $D^j f(x)$ the j^{th} differential of f at point x, and by $f^{[j]}(x,t)$ the quantity $\frac{1}{j!}D^j f(x,t)$; if f is analytic, the $f^{[j]}$'s are the coefficients of the power series.

We shall also need the following convention: if j and m are two strictly positive integers with $j \le m$, we shall denote by Σ_j^m the set of ordered sets $(l_1 \dots l_j)$ of j positive integers l_i, satisfying $\sum_{i=1}^{i=j} l_i = m$; given two sequences x_i and y_i, we shall denote, for any integer m and any $k \le m$

$$\Sigma_k^m(x_\bullet, y_\bullet) = \sum_{j=k}^{j=m} \sum_{(l_1 \dots l_j) \in \Sigma_j^m} x_j y_{l_1} \dots y_{l_j} \tag{36}$$

We recall the following result:

Proposition 1 (Derivatives of $g \circ f$) *Let g and f two functions of class $\mathcal{C}^r$, with f from $\mathbb{R}$ into $\mathbb{R}^n$, g from $\mathbb{R}^n$ into $\mathbb{R}^p$. Then $h = g \circ f$ is of class $\mathcal{C}^r$, with*

$$h^{[m]}(\varepsilon) = \Sigma_1^m \left(g^{[\bullet]}(f(\varepsilon)), f^{[\bullet]}(\varepsilon) \right) \tag{37}$$

$$= g^{[1]}(f(\varepsilon)) f^{[m]}(\varepsilon) + \Sigma_2^m \left(g^{[\bullet]}(f(\varepsilon)), f^{[\bullet]}(\varepsilon) \right) \tag{38}$$

Comments and notations Remark that, in the previous sum, $f^{[m]}$ appears only once under the form $g^{[1]}(f(\varepsilon))f^{[m]}(\varepsilon)$, that is, the coefficient of $f^{[m]}$ in $h^{[m]}$ is the first derivative of g taken at point $f(\varepsilon)$. Also, $f^{[0]}$ is present only as an argument of $g^{[j]}$, e.g. $g^{[j]}(f(\varepsilon))$. We shall extend the previous formula to the case $m = 0$ by setting

$$\Sigma_1^0 \left(g^{[\bullet]}(f(\varepsilon)), f^{[\bullet]}(\varepsilon) \right) = g(f(\varepsilon)) \tag{39}$$

5.1.2 Differentiation of z and x

When considering the m^{th} derivative of x and z with respect to ε, we shall assume that f and $\bar{f}$ are of class $\mathcal{C}^r$ with respect to x, $r \ge m$, and that Δ is of class $\mathcal{C}^{r+1}$. Standards results show that x and z are then of class $\mathcal{C}^r$ with respect to ε, as long as the solutions don't explode. We have

$$\frac{\partial}{\partial t}\left(\frac{1}{m!}\frac{\partial^m}{\partial \varepsilon^m}\right) z(\varepsilon,t) = \frac{1}{m!}\frac{\partial^m}{\partial \varepsilon^m}\left(\frac{\partial}{\partial t}\right) z(\varepsilon,t) \tag{40}$$

Notice that the computation of $z^{[m]}$ does not require the knowledge of the dependence w/r to t of the vector field that defines $\frac{dz}{dt}$. This is why we shall omit to mention t in the expression of $\bar{f}$, f, *and so on...* Remember nonetheless than the symbol $f^{[j]}$ will denote differentiation with respect to variables *other* than time t.

Straightforward computation gives now the following result:

Lemma 5.1 (Derivatives of z with respect to ε) *We shall denote by g the function $\frac{\partial \Delta}{\partial x} f$. We have:*

$$\begin{aligned}
\frac{dz^{[0]}}{dt} &= \bar{f}(x^{[0]}) - \varepsilon g(x^{[0]}) \\
\frac{dz^{[1]}}{dt} &= \bar{f}^{[1]}(x^{[0]})x^{[1]} - g(z^{[0]}) - \varepsilon g^{[1]}(x^{[0]})x^{[1]} \\
\frac{dz^{[m]}}{dt} &= \Sigma_1^m(f^{[\bullet]}(x), x^{[\bullet]}) - \varepsilon \Sigma_1^m(g^{[\bullet]}(x), x^{[\bullet]}) - \Sigma_1^{m-1}(g^{[\bullet]}(x), x^{[\bullet]})
\end{aligned}$$

with $z^{[0]}(t=0) = x_0$ and $z^{[m]}(t=0) = 0$ for $m \geq 1$.

Proof: We have $z^{[m]}(0) = 0$ for $m \geq 1$ since the initial condition of z does not depend on ε; the rest follows from the differentiation formula above. ∎

Let us compute now the derivatives of x:

Lemma 5.2 (Derivatives of x with respect to ε) *As before, we shall omit to mention t in the following expression. We have:*

$$x^{[0]} = z^{[0]} + \varepsilon \Delta(x^{[0]}) \tag{41}$$

$$x^{[1]} = z^{[1]} + \varepsilon \Delta^{[1]}(x^{[0]})x^{[1]} + \Delta(x^{[0]}) \tag{42}$$

$$x^{[m]} = z^{[m]} + \varepsilon \Sigma_1^m(\Delta^{[\bullet]}(x), x^{[\bullet]}) + \Sigma_1^{m-1}(\Delta^{[\bullet]}(x), x^{[\bullet]}) \tag{43}$$

Proof: We have

$$z(\varepsilon) = x(\varepsilon) - \varepsilon \Delta(x(\varepsilon)) \tag{44}$$

The rest follows from composite differentiation. ∎

Since we mean to estimate the $z^{[m]}$'s and $x^{[m]}$'s, we shall need an explicit (strictly causal) recursion on the $z^{[m]}$ and $x^{[m]}$.

Theorem 5.1 (Explicit recursion on the x_m) *Let $A(\varepsilon) = Id - \varepsilon \Delta^{[1]}(x^{[0]})$*[7]*. Then*

$$\begin{aligned}
x^{[1]} &= A^{-1}(\varepsilon)\left[z^{[1]} + \Delta(x^{[0]})\right] \\
x^{[m]} &= A^{-1}(\varepsilon)\left[z^{[m]} + \varepsilon \Sigma_2^m(\Delta^{[\bullet]}(x), x^{[\bullet]}) + \Sigma_1^{m-1}(\Delta^{[\bullet]}(x), x^{[\bullet]})\right]
\end{aligned}$$

where the right handside of the second line depends only on $z^{[m]}$ and $x^{[0]} \dots x^{[m-1]}$

[7] A is invertible for $\varepsilon \in [0,1]$ and x close to x_0

Proof: We have $\Sigma_1^m(\Delta^{[\bullet]}(x), x^{[\bullet]}) = \Delta^{[1]}(x^{[0]})x^{[m]} + \Sigma_2^m(\Delta^{[\bullet]}(x), x^{[\bullet]})$ where the last term depends only on $x^{[0]} \ldots x^{[m-1]}$; this gives

$$(Id - \varepsilon\Delta^{[1]}(x^{[0]}))x^{[m]} = z^{[m]} + \varepsilon\Sigma_2^m(\Delta^{[\bullet]}(x), x^{[\bullet]}) + \Sigma_1^{m-1}(\Delta^{[\bullet]}(x), x^{[\bullet]}) \quad (45)$$

■

Theorem 5.2 (Explicit recursion on the z_m) *Let $B(\varepsilon) = \bar{f}^{[1]}(x^{[0]}) - \varepsilon g^{[1]}(x^{[0]})$, $C(\varepsilon) = B(\varepsilon)A^{-1}(\varepsilon)$ (when defined), and Φ_ε the transition matrix of $C(\varepsilon)$. Then*

$$\begin{aligned} z^{[1]}(t) &= \int_0^t \Phi_\varepsilon(t,s)\left[C(\varepsilon)\Delta^{[0]}(x^{[0]}) - g^{[0]}(x^{[0]})\right]\, ds \\ z^{[m]}(t) &= \int_0^t \Phi_\varepsilon(t,s)\Big\{C(\varepsilon)\left[\varepsilon\Sigma_2^m(\Delta^{[\bullet]}(x), x^{[\bullet]}) + \Sigma_1^{m-1}(\Delta^{[\bullet]}(x), x^{[\bullet]})\right] \\ &\quad + \Sigma_2^m(\bar{f}^{[\bullet]}(x), x^{[\bullet]}) - \varepsilon\Sigma_2^m(g^{[\bullet]}(x), x^{[\bullet]}) - \Sigma_1^{m-1}(g^{[\bullet]}(x), x^{[\bullet]})\Big\}\, ds \end{aligned}$$

which, together with theorem 5.1, gives an explicit recursion on the $z^{[m]}$'s and $x^{[m]}$'s.

Proof: We have

$$\begin{aligned} \frac{dz^{[m]}}{dt} &= \bar{f}^{[1]}(x^{[0]})x^{[m]} + \Sigma_2^m(f^{[\bullet]}(x), x^{[\bullet]}) \\ &\quad -\varepsilon g^{[1]}(x^{[0]})x^{[m]} - \varepsilon\Sigma_2^m(g^{[\bullet]}(x), x^{[\bullet]}) \\ &\quad -\Sigma_1^{m-1}(g^{[\bullet]}(x), x^{[\bullet]}) \end{aligned}$$

The rest comes from the substitution of $x^{[m]}$ in the previous formula by the expression found in theorem 5.1. ■

Comment Since ε is an homotopy parameter between (almost) any two vectorfields, the previous expansions can be expected to be different from those found in the classical averaging literature ([6] [1]). Indeed, they are, since classical expansions rely on repeated averaging to be computed; this is not the case here, as we perform multiple integration instead. However, we shall see in the next section that the homotopy expansions still provide expansions with respect to the period in the periodic case (and similarly, for KBM vectorfields), and also with respect to the window length d in the case of windowed averaging.

5.2 Estimates

We assume that f and $\bar{f}$ are of class $\mathcal{C}^r$, and that Δ is of class $\mathcal{C}^{r+1}$. Let $\bar{F}_j$, G_j, D_j such that, for $j \geq 0$, and (x,t) in some domain V, we have

$$\|\bar{f}^{[j]}(x,t)\alpha_1 \ldots \alpha_j\| \leq \bar{F}_j\|\alpha_1\| \ldots \|\alpha_j\| \quad (46)$$

$$\|g^{[j]}(x,t)\alpha_1 \ldots \alpha_j\| \leq G_j\|\alpha_1\| \ldots \|\alpha_j\| \quad (47)$$

$$\|\Delta^{[j]}(x,t)\alpha_1 \ldots \alpha_j\| \leq D_j\|\alpha_1\| \ldots \|\alpha_j\| \quad (48)$$

for all $\alpha_1 \dots \alpha_j$, and j such that the above derivatives exist. We define $c(\varepsilon)$ and φ_ε as in section 3.2 (except we use G_1 directly for the time being).

Lemma 5.3 *Let ε be the point at which we differentiate x and z. Define the (positive) sequences $Z_m(\varepsilon)$ and $X_m(\varepsilon)$ by*

$$\begin{aligned}
Z_1(\varepsilon) &= \varphi_\varepsilon(c(\varepsilon)D_0 + G_0) \\
X_1(\varepsilon) &= \frac{Z_1(\varepsilon) + D_0}{1 - \varepsilon D_1} \quad \text{and, for } m > 1: \\
Z_m(\varepsilon) &= \varphi_\varepsilon \left\{ c(\varepsilon) \left[\varepsilon \Sigma_2^m(D_\bullet, X_\bullet(\varepsilon)) + \Sigma_1^{m-1}(D_\bullet, X_\bullet(\varepsilon)) \right] \right. \\
&\quad \left. + \Sigma_2^m(\bar{F}_\bullet, X_\bullet(\varepsilon)) + \varepsilon \Sigma_2^m(G_\bullet, X_\bullet(\varepsilon)) + \Sigma_1^{m-1}(G_\bullet, X_\bullet(\varepsilon)) \right\} \\
X_m(\varepsilon) &= \frac{1}{1 - \varepsilon D_1} \left[Z_m(\varepsilon) + \varepsilon \Sigma_2^m(D_\bullet, X_\bullet(\varepsilon)) + \Sigma_1^{m-1}(D_\bullet, X_\bullet(\varepsilon)) \right]
\end{aligned}$$

Then, as long as $(x(\varepsilon_1, t), t)$ stays in V for all $\varepsilon_1 \in [0, \varepsilon]$ (by satisfying condition $\mathcal{V}_R$, for instance), we have $\|x^{[m]}(\varepsilon, t)\| \le X_m(\varepsilon)$ and $\|z^{[m]}(\varepsilon, t)\| \le Z_m(\varepsilon)$ for $1 \le m \le r$, and Z_m and X_m are non decreasing functions of ε.

Proof: we use the estimates of lemma 3.1. The rest follows from recursion and the formulas of theorems 5.1 and 5.2; notice that the expressions in the right hand-sides of this theorem are all non decreasing with respect to all of their arguments. ■

The trouble is that we do not know yet what the sequences X_m and Z_m look like. We are going to see that, under certain assumptions, they can be interpreted as the derivatives of some functions X and Z. To achieve this, we restrict our attention to the sequences $X_m(1)$ and $Z_m(1)$, since they provide estimates of $x^{[m]}(\varepsilon)$ and $z^{[m]}(\varepsilon)$ for all ε in [0,1].

Lemma 5.4 *Let X_0 some positive scalar, $\bar{F}$ (resp. D, G) a regular scalar function such that $\bar{F}^{[m]}(X_0) = \bar{F}_m$ (resp. $D^{[m]}(X_0) = D_m$, $G^{[m]}(X_0) = G_m$). Let $X(\varepsilon)$ and $Z(\varepsilon)$ satisfy*

$$\begin{aligned}
Z + \varphi_1 \left(c(1)D_1 + \bar{F}_1 + G_1 \right) X &= \varphi_1 \left(c(1)\varepsilon D(X) + \bar{F}(X) + \varepsilon G(X) \right) && (49) \\
X &= Z + \varepsilon D(X) && (50)
\end{aligned}$$

for ε in a neighborhood in 1. Then, if $X(1) = X_0$, we have $X_m(1) = X^{[m]}(1)$ and $Z_m(1) = Z^{[m]}(1)$ for $m \ge 1$, that is, the estimates of the derivatives are the derivatives of some functions at $\varepsilon = 1$.

Proof: This is obtained by differentiating relations (49) and (50) at $\varepsilon = 1$, and checking the expressions against those of the previous lemma **at** $\varepsilon = 1$. ■

Comment Having $X^{[m]}(\varepsilon) = X_m(\varepsilon)$ and $z^{[m]}(\varepsilon) = Z_m(\varepsilon)$ for all ε is not compatible with equations (49) and (50); this is why we have to limit ourselves to the case $\varepsilon = 1$.

Theorem 5.3 (Estimates in the $\mathcal{C}^r$ case) *Let $\bar{F}$, G, D and ρ four positive numbers, with $\rho < 1$, such that $\bar{F}_j \leq \bar{F}/\rho^j$, $G_j \leq G/\rho^j$ and $D_j \leq D/\rho^j$. We assume that $D \leq \rho$, which is equivalent to having $D_1 \leq 1$. Let $\alpha = \varphi_1(c(1)D + G)$, $\beta = \bar{F}_1\varphi_1$. Let σ_k the sequence such that $1 - \sqrt{1-x} = \sum_{k=1}^{\infty} \sigma_k x^k$. Then, as long as $(x(\varepsilon, t), t)$ stays in V for all $\varepsilon \in [0, 1]$ (by satisfying condition $\mathcal{V}_R$, for instance), one has*

$$\|x(1,t) - \sum_{m=0}^{m=r-1} x^{[m]}(0,t)\| \leq \sigma_r \frac{(D+\alpha)^r}{(\frac{\rho-D}{2})^{2r-1}} \tag{51}$$

Proof: Considering the assumptions, we can take

$$D(X) = \frac{D}{1 - \frac{X-X_0}{\rho}} \ , \ \bar{F}(X) = \frac{\bar{F}}{1 - \frac{X-X_0}{\rho}} \ , \ G(X) = \frac{G}{1 - \frac{X-X_0}{\rho}} \ , \tag{52}$$

with $X(\varepsilon)$ and X_0 that remain to be determined in order to satisfy the assumptions of the previous lemma. After eliminating Z from equations (49) and (50), we get:

$$X - \varepsilon \frac{D}{1 - \frac{X-X_0}{\rho}} + \varphi_1\left(c(1)D + \bar{F} + G\right)\frac{X}{\rho} = \frac{\varphi_1\left(c(1)\varepsilon D + \bar{F} + \varepsilon G\right)}{1 - \frac{X-X_0}{\rho}} \tag{53}$$

that is, if we let $Y = (X - X_0)/\rho$

$$(X_0 + \rho Y) - \varepsilon \frac{D}{1-Y} + \frac{\alpha+\beta}{\rho}(X_0 + \rho Y) = \frac{\beta + \varepsilon\alpha}{(1-Y)} \tag{54}$$

For $\varepsilon = 1$, we must have $Y = 0$, that is:

$$X_0 = \frac{\rho(D+\alpha+\beta)}{\rho+\alpha+\beta} \tag{55}$$

Equation then (54) becomes

$$Y^2 - Y\frac{D-\rho}{\rho+\alpha+\beta} - \frac{(1-\varepsilon)(\alpha+D)}{\rho+\alpha+\beta} = 0 \tag{56}$$

The discriminant is positive; there is only one positive root, which gives:

$$Y(\varepsilon) = \frac{1}{2(\rho+\alpha+\beta)}\left[D - \rho + \sqrt{(\rho-D)^2 - 4(\varepsilon-1)(D+\alpha)}\right] \tag{57}$$

We have $Y(1) = 0$ if and only if $D \leq \rho$.

If we denote by $1 - \sum_1^{\infty} \sigma_k x^k$ the power series associated to the analytical function $\sqrt{1-x}$, we can derive, thanks to the preceding lemma, the following expression for X_r:

$$X_r = \sigma_r \frac{\rho}{\rho+\alpha+\beta}\frac{(D+\alpha)^r}{(\frac{\rho-D}{2})^{2r-1}} \leq \sigma_r \frac{(D+\alpha)^r}{(\frac{\rho-D}{2})^{2r-1}} \tag{58}$$

The result follows now from the Taylor formula:

$$x(1,t) = \sum_{m=0}^{m=r-1} x^{[m]}(0,t) + x^{[r]}(\theta,t) \quad , \quad \theta \in]0,1[\tag{59}$$

and from lemma 5.3. ■

5.3 Discussion on the estimates

In this section, we are going to prove that the estimation of the rest in theorem 5.3 is essentially of the form $(D/\rho^2)^r$. We will use then this result to prove that the expansions of the previous sections are also expansions with respect to the period μ in the periodic case, or in the case of KBM vectorfields, and with respect to the window width d in the case of windowed averaging (the previous being actually a particular case of the latter).

By looking at (51), we see that our claim will be true if the composite variable α can be bound linearly with respect to D. Let us first show that G can be bound linearly with respect to D, at least for a finite number of differentiations.

Lemma 5.5 *Assume that Δ satisfies $\|\Delta^{[k]}\| \leq D/\rho^k$ and that f satisfies $\|f^{[k]}\| \leq F/\rho^k$, for $0 \leq k \leq r$. Define $\Gamma = \frac{F}{\rho}\frac{(r+2)(r+1)}{2}$. Then $g = \Delta^{[1]} f$ satisfies*

$$\|g^{[k]}\| \leq \frac{\Gamma D}{\rho^k} \tag{60}$$

Proof: We have, from the definition of g:

$$\|g^{[k]}\| \leq \left(\frac{DF}{\rho\left(1 - \frac{X}{\rho}\right)^3} \right)^{[k]}_{(X=0)} \tag{61}$$

Since $(1 - x/\rho)^{-3}$ is equal to $\rho^2/2$ multiplied by the second derivative of $(1 - x/\rho)^{-1}$, we see that the second hand of the preceding inequality is equal to $\frac{DF(k+2)(k+1)}{2\rho^{k+1}}$. This quantity is smaller than some G/ρ^k if we take, for instance,

$$G = \max_k \frac{DF(k+2)(k+1)}{2\rho} = \frac{DF(r+2)(r+1)}{2\rho} \stackrel{\text{def}}{=} \Gamma D \tag{62}$$

■

Let us us draw now a finer estimation on the rest of the Taylor expansion, independently of the averaging method used. After that, we shall particularize this result to the case of windowed averaging, periodic averaging, etc…

Theorem 5.4 *Assume that D is such that*

$$D \leq \frac{\rho}{2} \min\left(1, \frac{1}{2\varphi_0(\bar{F}/\rho + \Gamma)}\right) \tag{63}$$

Then

$$\|x(1,t) - \sum_{m=0}^{m=r-1} x^{[m]}(0,t)\| \leq \sigma_r D^r \frac{(1+2\varphi_0(\bar{F}/\rho+\Gamma))^r}{(\frac{\rho}{4})^{2r-1}} \tag{64}$$

Proof: Assumption (63) is there just to make sure that $D \leq \rho/2$ and $c(1)\varphi_0 \leq 1/2$. We see that we have then $\varphi_1 \leq 2\varphi_0$, and that $c(1) \leq \bar{F}/\rho + \Gamma$. This leads to

$$\alpha \leq 2D\varphi_0(\frac{\bar{F}}{\rho} + 2\Gamma) \tag{65}$$

The rest follows from theorem 5.3. ∎

To get asymptotic estimates, we have now to make distinctions on the different possible averaging schemes. One good reason lies in the fact that, depending on the methods, the "small" averaging parameter defines a family of averages $\bar{f}$ (such as for d in the windowed averaging scheme[8]), or a family of "original" vectorfields f (such as in periodic averaging, or KBM dynamics). Let us start with the windowed averaging method.

5.3.1 The windowed averaging case

Let us first notice that the derivatives (with respect to x) of the windowed average $\bar{f}$ are also the windowed averages of the corresponding derivatives of f. This implies that, if f satisfies $\|f^{[j]}\| \leq F/\rho^j$, then the same Cauchy estimates can be drawn on $\bar{f}$, e.g. $\|\bar{f}^{[j]}\| \leq F/\rho^j$, whatever the choice of the width d. Moreover, we can draw from theorem 4.1 the following result:

Theorem 5.5 *Assume that f satisfies $\|f^{[k]}\| \leq F/\rho^k$ on a neighborhood $\mathcal{V}$ of the initial conditions, and consider Δ as defined by the windowed averaging scheme of section 4. Then Δ satisfies*

$$\|\Delta^{[k]}(x,\bullet)\|_\infty \leq d\sigma(f^{[k]}) \leq d\frac{F}{\rho^k} \tag{66}$$

Proof: This comes straight from theorem 4.1. ∎

The conclusion from this is that we can take $\bar{F} = F$ and $D = dF$, and that ρ and Γ can be chosen independently of d. However, φ_0 depends on d, since the average itself depends on d. Thanks to lemma 3.2, we know that $\varphi_0 \leq \Phi_1/(1 - c(1)\Phi_1)$ where Φ_1 denotes the convolution norm of the transition matrix at $\varepsilon = 1$ (which is different from φ_1); Φ_1 is independent of d.

We use all this to get the main result:

Theorem 5.6 *Let d such that*

$$d \leq \frac{\rho}{2F} \min\left(1, \frac{1}{4\Phi_1(F/\rho+\Gamma)}\right) \tag{67}$$

[8] It is interesting to remark that, in this case, the "limit" average (e.g., when the small parameter goes to 0) is the original function itself. This means that gains are to be expected from averaging when values of d that are neither too small, nor too large.

Denote by $\bar{f}_d$ the windowed average of f with step d and $x_d(\varepsilon, t)$ the trajectories defined by the averaging homotopy between the dynamics f and $\bar{f}_d$, and Φ_1 the convolution norm of the tangent system of f at x_0. Note that $x_d(1,t)$ does not depend on d. Then

$$\|x(1,t) - \sum_{m=0}^{m=r-1} x_d^{[m]}(0,t)\| \leq \sigma_r(dF)^r \frac{(1+4\Phi_1(F/\rho+\Gamma))^r}{(\frac{\rho}{4})^{2r-1}} \tag{68}$$

and the homotopy expansion is also an asymptotic expansion with respect to d.

Proof: We take $D = dF$. The assumption on d makes sure that $c(1)$ is smaller than $1/2\Phi_1$; then φ_0 is smaller than $2\Phi_1$, and we retrieve assumption (63). The rest follows from theorem 5.4. ∎

5.3.2 KBM vectorfields

In opposition to the previous case, we have here a unique average and a multiplicity of non averaged dynamics. These are under the form $f_\mu(x,t) = f(x,t,t/\mu)$. We shall assume that $f(x,t,\tau)$ satisfies some ergodicity assumption in τ, and that Cauchy estimates on its derivatives with respect to x and t can be drawn uniformly in τ, something that is quite reasonable in the periodic case. We shall eliminate the dependence of f_μ and $\bar{f}$ in t by augmenting the two dynamics with $dt/dt = 1$. Then the estimates on the new dynamics include the behavior of the original ones with respect to t. The corresponding Δ function is the original augmented with a zero. Since the dynamics f_μ are actually expressed in terms of a fixed functional f, it is natural to assume that the Cauchy estimates F/ρ^k and $\bar{F}/\rho^k$ are known from the start; considering the assumptions above, these estimates can be drawn independently of τ. This means in particular that F, $\bar{F}$ and ρ can be computed independently of the small parameter μ. The same can be said of φ_0 and Γ.

The ergodicity assumption will be the following: there exists a bound M such that

$$\left| \int_0^T (f^{[k]}(x,\tau) - \bar{f}^{[k]}(x))d\tau \right| \leq \frac{M}{\rho^k} \ , \ 0 \leq k \leq r \tag{69}$$

for all positive T and all x in the domain $\mathcal{V}$ where the other estimates are drawn.

This means that the limit in (3) goes to 0 like $1/T$. This is a reasonable assumption that is satisfied in the periodic case. In fact, the derivatives of the averages are, in the case of periodic as well as windowed averaging, the averages of the derivatives; this means that we can always take $M = 2F$ in these cases. We have the following result

Lemma 5.6 *Define $\Delta_\mu(x,t) = \int_0^t (\bar{f}(x) - f_\mu(x,s))ds$. Then Δ_μ satisfies*

$$\|\Delta_\mu^{[k]}(x,t)\| \leq \mu \frac{M}{\rho^k} \tag{70}$$

Proof: A simple change of integration variable and the use of assumption (69) do the trick. ■

We can now go to the main result:

Theorem 5.7 *Let μ such that*

$$\mu \leq \frac{\rho}{2M} \min\left(1, \frac{1}{2\varphi_0(\bar{F}/\rho + \Gamma)}\right) \tag{71}$$

Denote by $x_\mu(\varepsilon, t)$ the trajectories defined by the averaging homotopy between the dynamics f_μ and $\bar{f}$. Note that $x_\mu(0,t)$ does not depend on μ. Then

$$\|x_\mu(1,t) - \sum_{m=0}^{m=r-1} x^{[m]}(0,t)\| \leq \sigma_r (\mu M)^r \frac{(1 + 2\varphi_0(\bar{F}/\rho + \Gamma))^r}{(\frac{\rho}{4})^{2r-1}} \tag{72}$$

and the homotopy expansion is also an asymptotic expansion with respect to μ.

Proof: We can take $D = \mu M$. The rest follows from theorem 5.4. ■

Comment It is interesting to notice that, depending on the case (windowed or classical averaging), the limit lies at one end of the homotopy or at the other. This shows that there is no reason to privilege one endpoint at the expense of the other, and, at least in the opinion of the author, the expression of averaging in terms of a distance (Δ in this case) between two dynamics is more flexible than the asymptotic approach.

6 Conclusion

We have show that the classical averaging theory which is base on time scales in the dependency of the dynamics with respect to time can be reformulated under an homotopty between two dynamical systems. This homotopy is similar to regular perturbations in the sense that a small difference in *amplitude* between the two dynamics leads to close trajectories. It also encompasses classical averaging, with the difference that the perturbed dynamics $\bar{f}$ is to required to be defined by some ergodic approximation of f, but can be freely defined by the user; typically, signal processing tools can be used to define $\bar(f)$ for the time behavior of f. First order approximations and higher order approximations have been studied and the convergence of the series for $\epsilon = 1$ has been investigated.

The next step is the generalization of the extension of this theory to input/output systems and should the subject of future work.

References

[1] S.N. Chow and J.K. Hale. *Methods of Bifurcation.* Springer-Verlag, 1982.

[2] Ingrid Daubechies. *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, 1992.

[3] W. Eckhaus. New approach to the asymptotic theory of nonlinear oscillations and wave-propagation. *J. of Math. Analysis and App.*, 49:575–611, 1975.

[4] Stéphane Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1998.

[5] J.A. Sanders and F. Verhulst. *Averaging Methods in Nonlinear Dynamical Systems*. Springer-Verlag, 1985.

[6] V.M. Volosov. Averaging in systems of ordinary differential equations. *Russ. Math. Surveys*, 17:1–128.